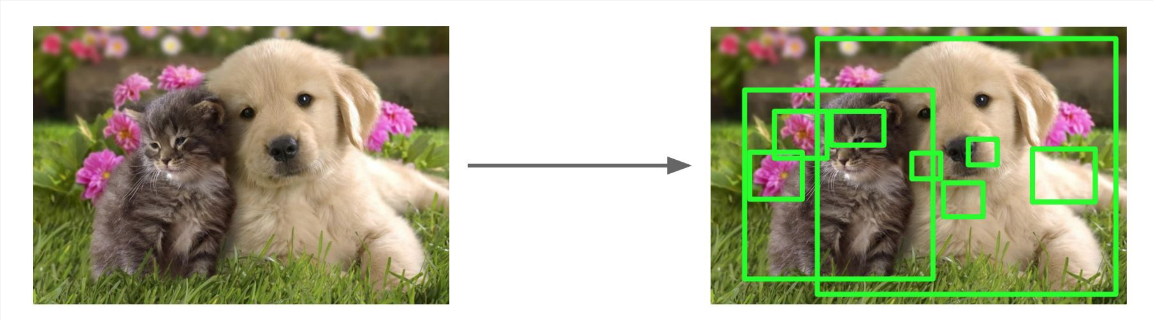


Graphical Models in Detecting Regions of Interest (RoI)

ECE/CS 8803 - Probabilistic Graphical Models

Jason Lin, Hakki Torun, Weihua Zhu, Jingjing Pan

First RoI Measure: Objectness

- **"Region of Interest (RoI)"**
  - A portion of image that you are interested and want to perform some other operation on

First RoI measure: Objectness

- **“Region of Interest (RoI)”**
  - A portion of image that you are interested and want to perform some other operation on
- **“Objectness”**
  - Measures how likely an image window contains an object of any class
  - Can be useful as low-level preprocessing

First RoI Measure: Objectness

- **"Region of Interest (RoI)"**
  - A portion of image that you are interested and want to perform some other operation on
- **"Objectness"**
  - Measures how likely an image window contains an object of any class
  - Can be useful as low-level preprocessing
- **Our motivation**
  - Use Gaussian Process based Bayesian Optimization to find better parameters for objectness algorithm
  - Detect object regions better in the difficult images

Objectness Algorithm: ***What is an object?*** **[Alexe+, CVPR 2010][1]**

- Combine in a Bayesian framework 4 image cues measuring characteristics of objects:
  - **MS: Multi-scale Saliency** *(5 parameters)*
  - CC: Color Contrast *(1 parameter)*
  - ED: Edge Density near window borders *(1 parameter)*
  - SS: Superpixel Straddling *(1 parameter)*

Objectness Algorithm: *What is an object?* [Alexe+, CVPR 2010]

- 50 training images
- 100,000 random windows from each image
- Define windows with >0.5 IoU as positive windows (and <=0.5 as negative)
- For each of the 8 parameters, maximize posterior probability:

$$\theta^* = \arg\max_{\theta} \prod_{w \in \mathcal{W}^{\text{obj}}} p_\theta(\text{obj}|\text{CC}(w,\theta)) =$$
$$= \arg\max_{\theta} \prod_{w \in \mathcal{W}^{\text{obj}}} \frac{p_\theta(\text{CC}(w,\theta)|\text{obj}) \cdot p(\text{obj})}{\sum_{c \in \{\text{obj},\text{bg}\}} p_\theta(\text{CC}(w,\theta)|c) \cdot p(c)}$$

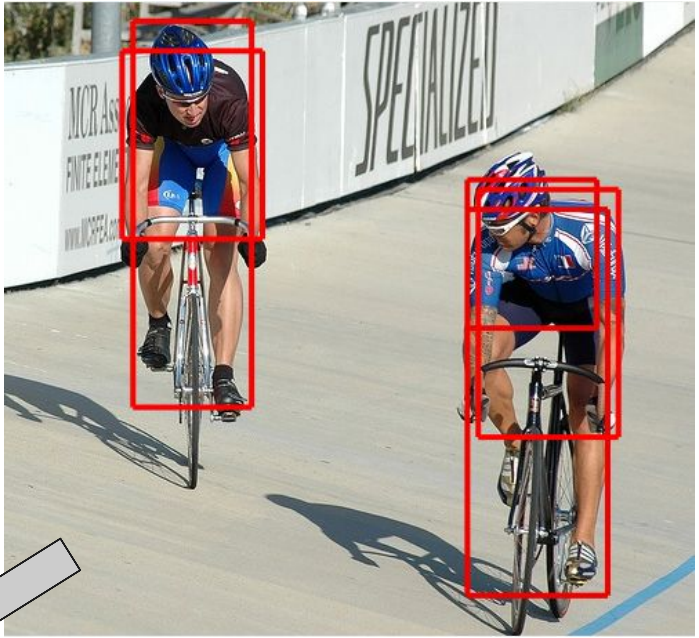
Model Output

- Generate arbitrary number of target proposal windows with the potential probability of containing objects.
- Arbitrary numbers of proposal boxes with ranking order with the probabilities

Model Output

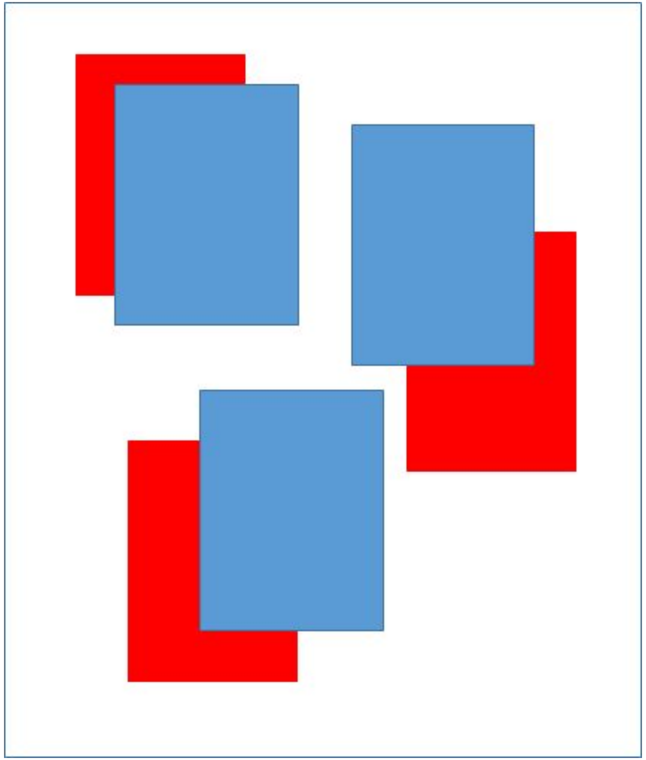
X1	Y1	X2	Y2	P1
X1	Y1	X2	Y2	P2
X1	Y1	X2	Y2	P3
X1	Y1	X2	Y2	P4
X1	Y1	X2	Y2	P5
X1	Y1	X2	Y2	Pi
X1	Y1	X2	Y2	Pi+1
X1	Y1	X2	Y2	Pi+2
.....	.....	.....	.....	.....
X1	Y1	X2	Y2	Pn

Top 5 Probability



343.7500	136.2656	445.3125	315.5625	0.9425
85.9375	35.8594	187.5000	172.1250	0.9369
93.7500	14.3438	179.6875	294.0469	0.9340
335.9375	150.6094	437.5000	430.3125	0.8896
335.9375	129.0938	429.6875	236.6719	0.8737

Objectness Evaluation



- Use mean Average Precision evaluation method for Top 5 windows

$$a_0 = \frac{Area(b_p \cap b_t)}{Area(b_p \cup b_t)}$$

$$w_i = p_i$$

$$score(\mathrm{CC,ED,SS,MS})$$

$$= \sum_{i=1}^{5} w_i \mathrm{I}(\max\{\frac{Area(p_i \cap p_{gt})}{Area(p_i \cup p_{gt})}\} > threshold)$$

Gaussian Process and Bayesian Optimization: Background

- Gaussian process defines a distribution over *functions*, which can be used to do Bayesian regression.

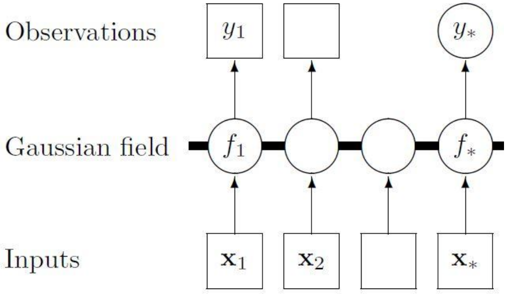
$$p(f|Data) = \frac{p(Data|f)p(f)}{p(D)}$$

- GPs are parametrized by their mean and covariance function

$$p(f) = \mathcal{N}(\mu(x), K)$$

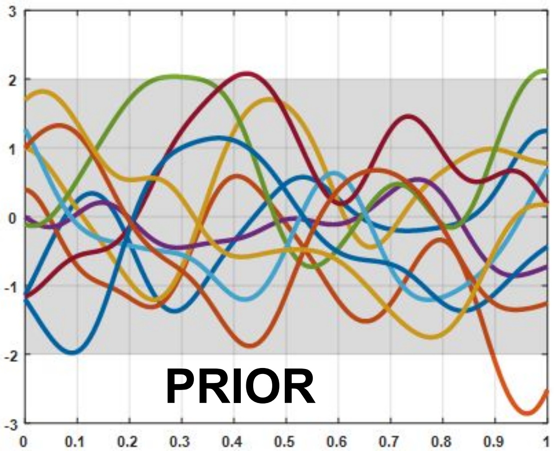
- After $\mu(x)$ and $k(x, x')$ is defined for prior, the conditional distribution of test points, (x^*, f^*), is defined as:

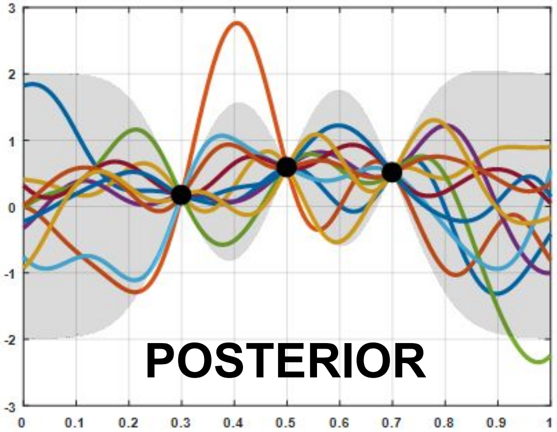
$$p(f^*|x^*, D) = \int p(f^*|x^*, f, D)p(f|D)df$$



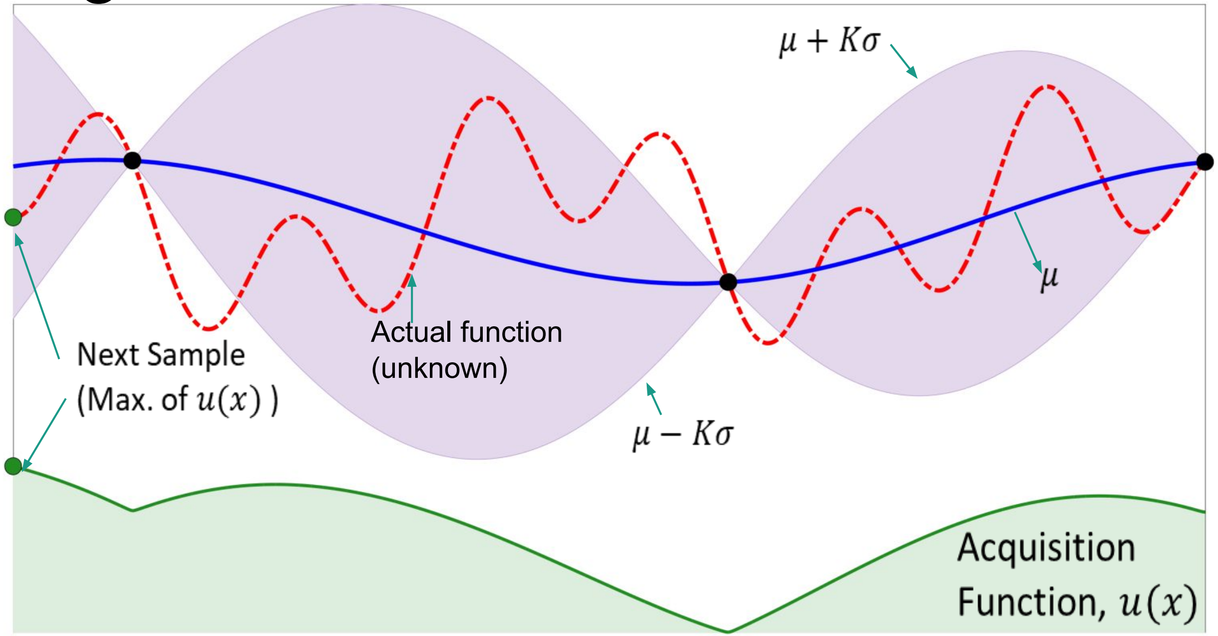
Using the conditional independence behavior defined in graph, the predictions can be computed as:

$$\mu(f^*|x^*, D) = k(x^*, x_D)^T K^{-1} y_D$$
$$\sigma(x_{t+1}) = \mathrm{k}(\mathrm{x}^*, \mathrm{x}^*) - k(x^*, x_D)^T K^{-1} k(x^*, x_D)$$





Gaussian Process and Bayesian Optimization: Background



- At each iteration, BO trains a <u>predictive Gaussian Process (GP)</u> using the observed data.
- The goal is <u>not to predict whole function</u>, but <u>to predict where global optima is</u>.
- Mean and variance of GP is used to create an <u>acquisition function, $u(x)$.</u>
- Next sampling point is determined by finding maximum of $u(x)$.(AUXILLARY OPTIMIZATION)

Predictive GP:

$$\mu(x_{t+1}) = k^T K^{-1} y_{1:t}$$
$$\sigma(x_{t+1}) = k(x_{t+1}, x_{t+1}) - k^T K^{-1} k$$

Kernel, k **&** K **:**

$$K_{1:t} = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_t) \\ \vdots & \ddots & \vdots \\ k(x_t, x_1) & \cdots & k(x_t, x_t) \end{bmatrix}$$

An Example Kernel Function:

ARD Matern 5/2:

$$k(x_i, x_j) = \sigma_f^2 \left(1 + \sqrt{5}r + \frac{5}{3}r^2\right) e^{-\sqrt{5}r}$$

$$r = \sqrt{\sum_{d=1}^{D} \frac{(x_{i,d} - x_{j,d})^2}{\sigma_d^2}}$$

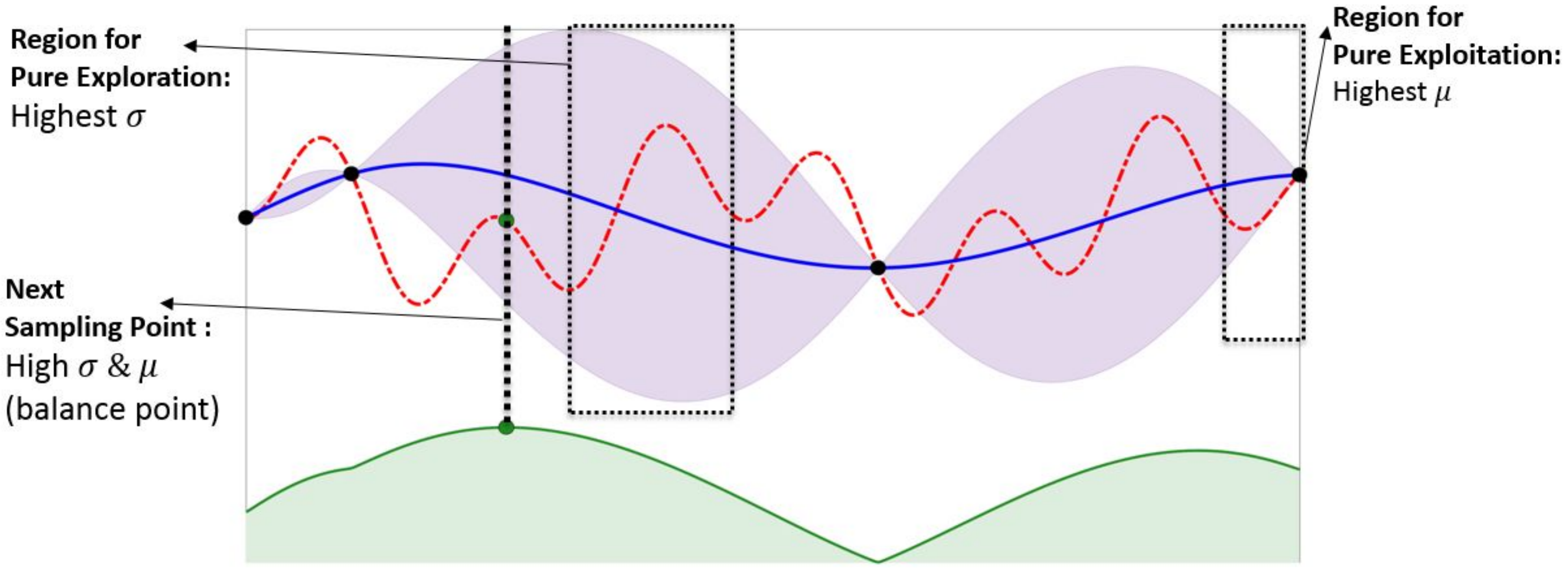
Training of GP:

Find $(\sigma_f, \sigma_1, \sigma_2, \ldots, \sigma_D)$ *such that marginal likelihood of GP,* L*, is max.*

$$L = \frac{1}{2}\log(\|C\|) + \frac{1}{2} y^T C^{-1} y$$

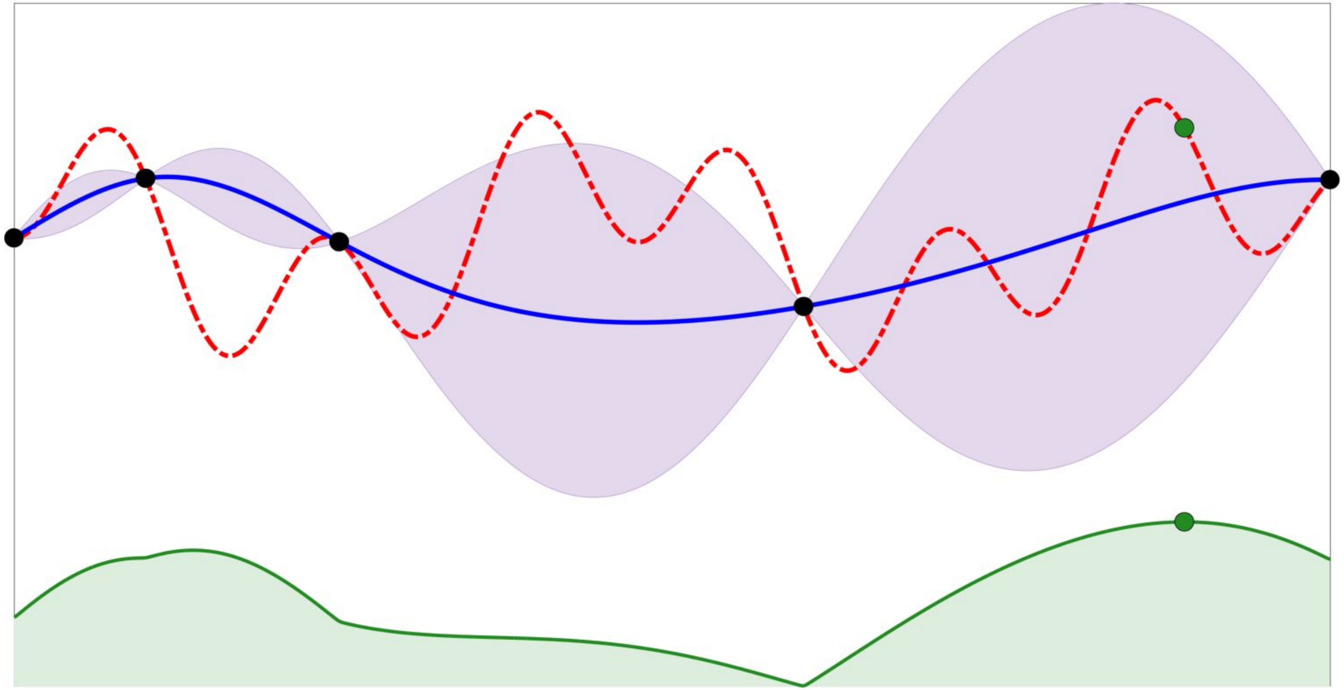
$C = K + \sigma^2 I, \quad \sigma \Rightarrow$observation noise

Gaussian Process and Bayesian Optimization: Background



- ❖Sampling strategy/acquisition function should balance exploration & exploitation.
  - ➢<u>Exploration:</u> Sampling in uncertain areas to learn the underlying function (high σ areas).
  - ➢<u>Exploitation:</u> Sampling around already good observation. (high μ areas).

Gaussian Process and Bayesian Optimization: Background



Most Popular Acquisition Function:

<u>Probability of Improvement</u>

Probability of increasing best observed value

$$u(x) = \Phi\left(\frac{\mu(x) - \widetilde{f^*} - \zeta}{\sigma(x)}\right)$$

<u>Expected Improvement</u>

The most probable point to give highest increment

$$u(x) = \sigma(x)[Z\Phi(Z) + \phi(Z)]$$

where $Z = \frac{\mu(x) - \widetilde{f^*} - \zeta}{\sigma(x)}$

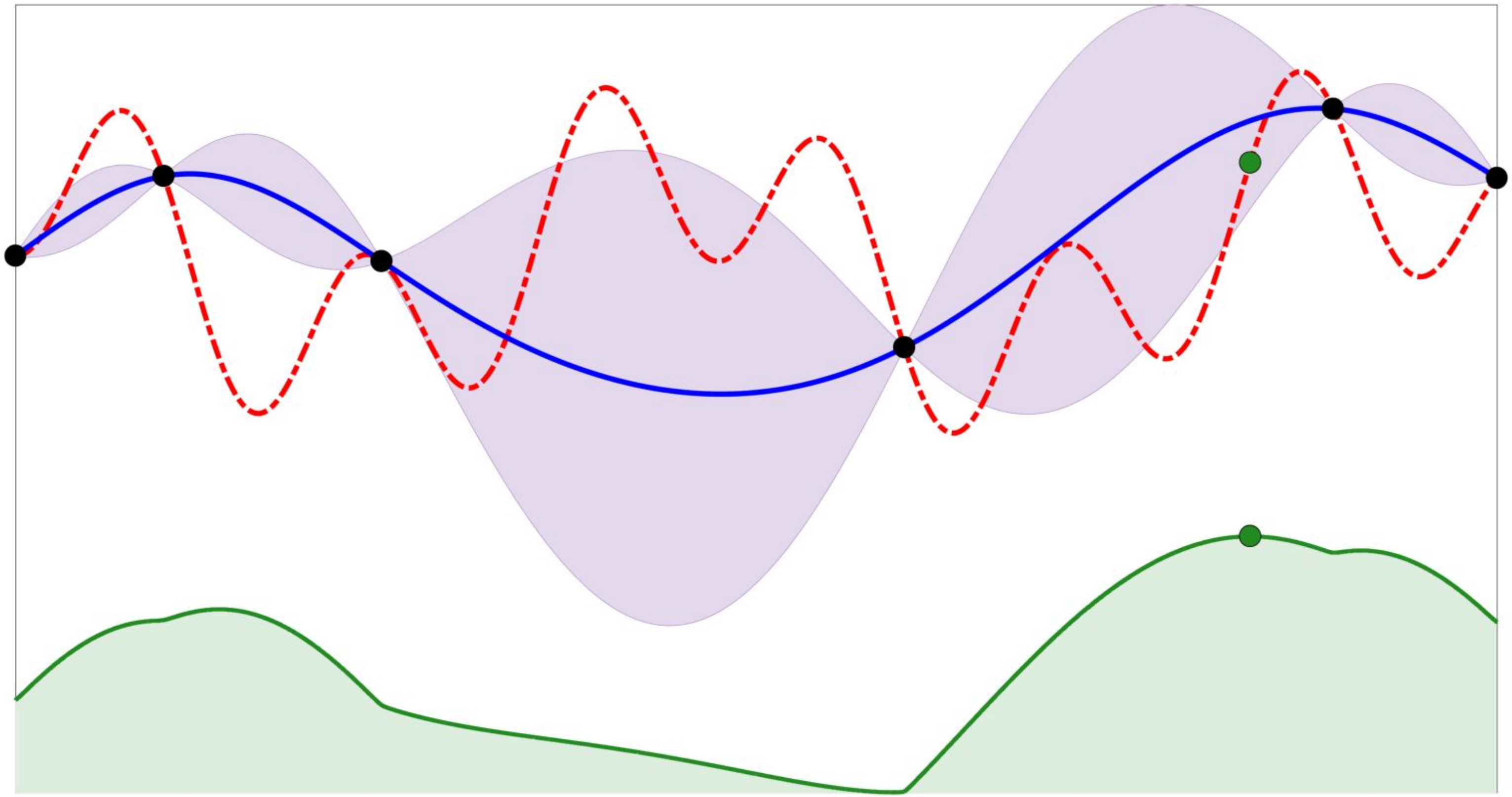
<u>Upper Confidence Bound:</u>

$$u(x) = \mu(x) + K\sigma(x)$$

where $K = \sqrt{2\ln(2\pi M^2/(12\eta)}$

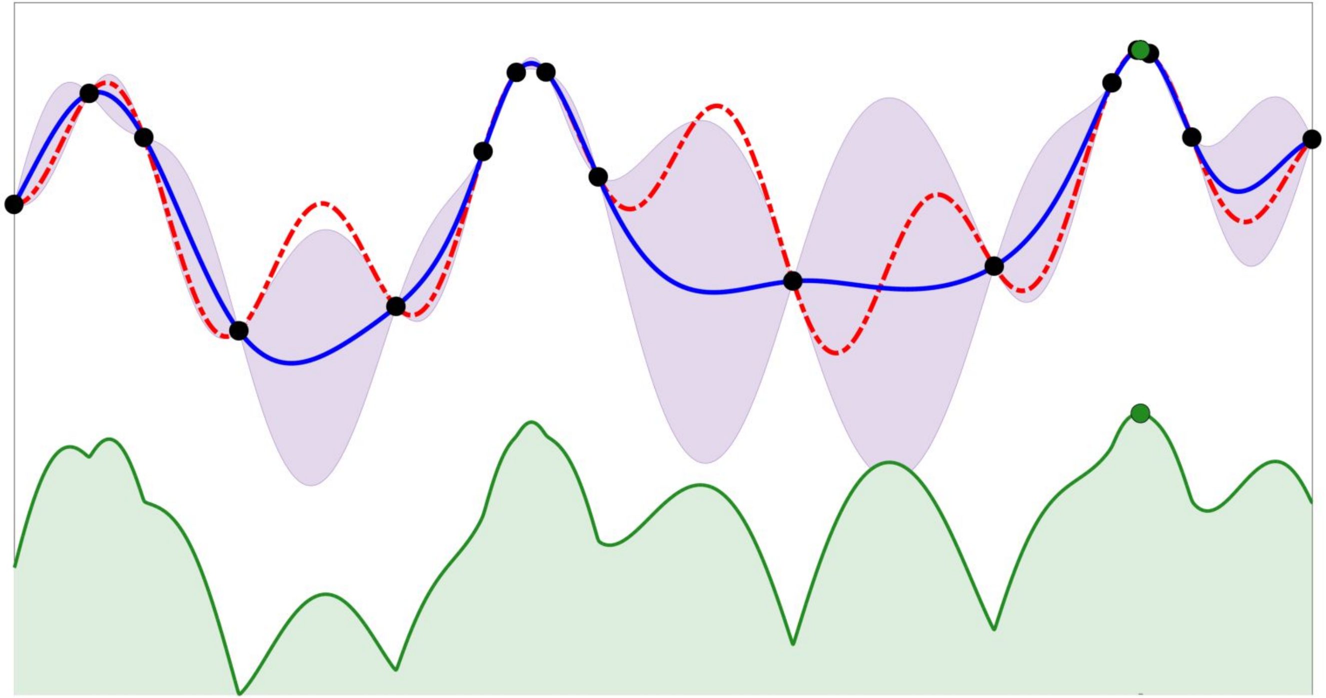
❖At each iteration, acquisition function selects the next point by balancing exploration & exploitation.

Gaussian Process and Bayesian Optimization: Background



❖The sampling point is chosen around global optima, as both μ & σ are high.

Gaussian Process and Bayesian Optimization: Background



Examples of stopping conditions

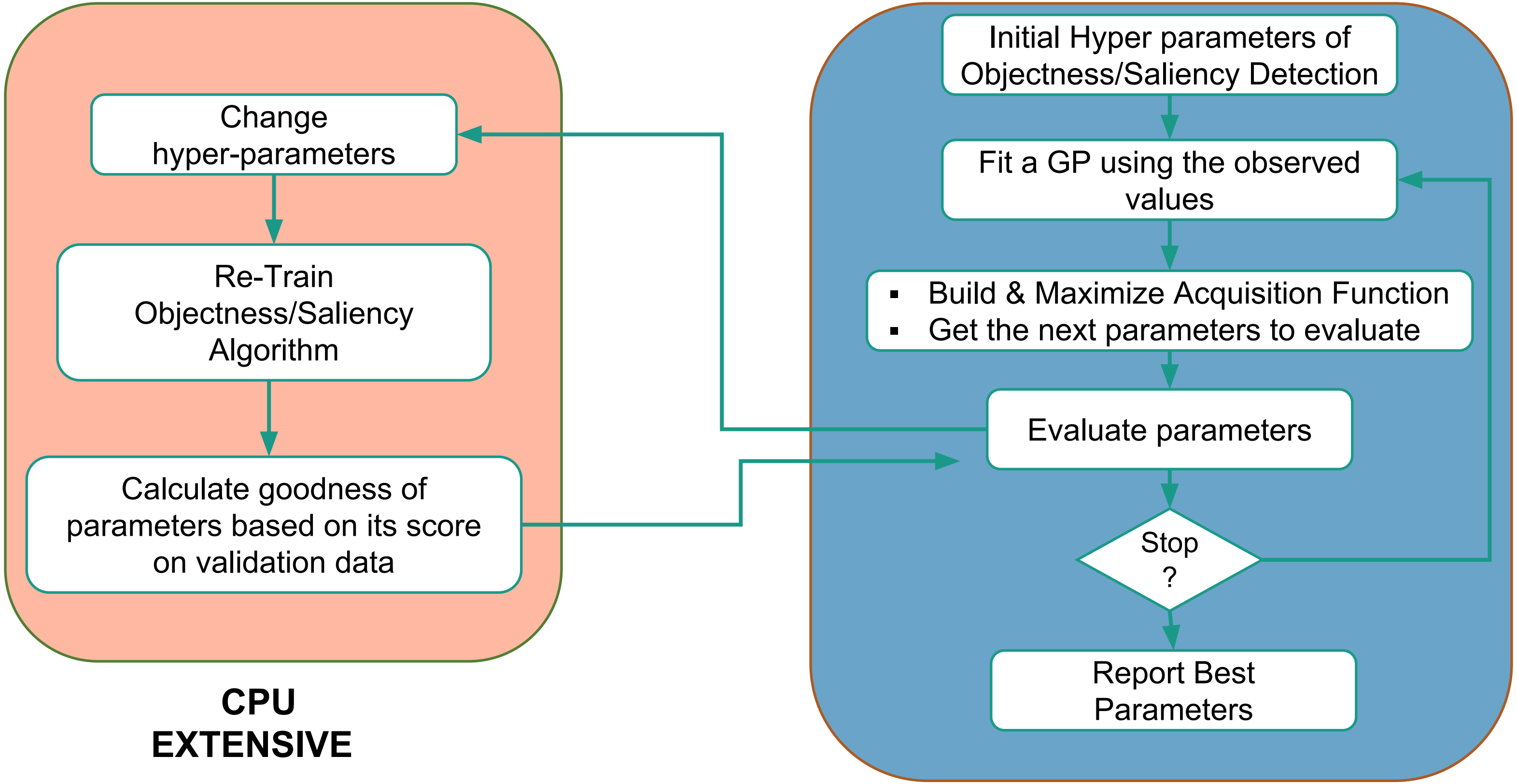
1) Highest UCB is lower than best observed point.
   (Global Optima w/ high probability)

2) Design Specification satisfied.

3) Maximum number of expensive simulations are exceeded.

❖After enough exploration, the previously found point is labelled as global optima since maximum UCB is lower than highest observed value
❖A proper stopping condition should be selected.

Optimization Framework to Apply BO on Objectness/Saliency Applications



Second RoI Measure: Saliency Detection

- **Problem Definition**:
  Selectively focus on parts of an image that distinguishes from surrounding features. Notion of relevance.
- **Two Paradigms**:
  Bottom-up (stimulus driven) vs. Top-down (goal driven)
- **Three stages:**
  S1 - feature extraction
  S2 - saliency activation
  S3 - master map normalization/combination
- **FIGRIM Fixation Dataset**:
  - 21 indoor/outdoor scene categories
  - 15 observers per image, 2s duration

Graph-based random walks: *"Graph-based Visual Saliency"* [Harel, NIPS 2006][4]

- **Motivation:**
  - Use Gaussian Process based Bayesian Optimization to find optimal weights for saliency maps combination algorithm (step 3)
  - Higher area under ROC curve w.r.t. ground truth annotations
- **Unified approach to steps 2-3** using <u>dissimilarity</u> and <u>saliency</u> as edge weights on Markov chains.
  - **Feature channels (7 parameters: CIORFMD)**
  - Final map gaussian blur fraction (1 param)
  - σ in activation, normalization step (2 params)
  - # normalization iterations (1 parameter)
  - Stopping threshold of evaluating equilibrium distribution (1 param)

$$d((i,j)||(p,q)) \triangleq \left|\log \frac{M(i,j)}{M(p,q)}\right|$$

$$w_1((i,j),(p,q)) \triangleq d((i,j)||(p,q)) \cdot F(i-p, j-q),$$

$$F(a,b) \triangleq \exp\left(-\frac{a^2+b^2}{2\sigma^2}\right).$$

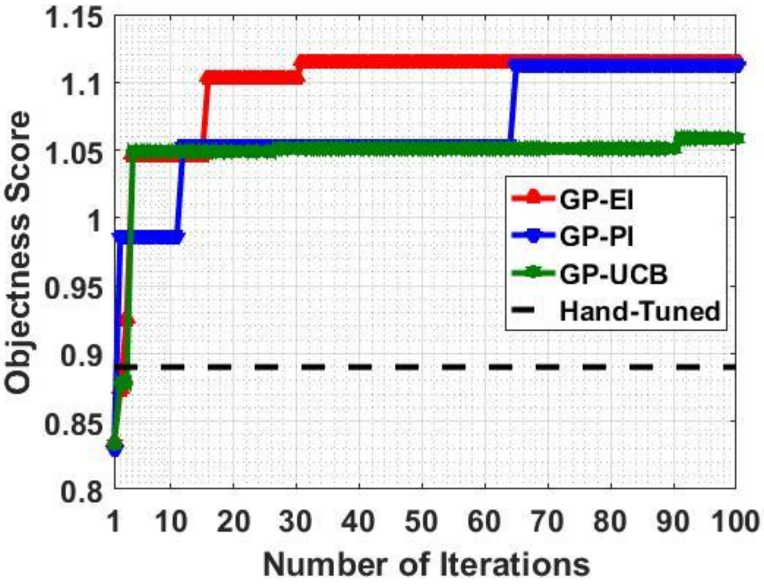
Supervised learning using Conditional Random Fields:

"Learning to Detect A Salient Object [Liu, CVPR 2007][5]

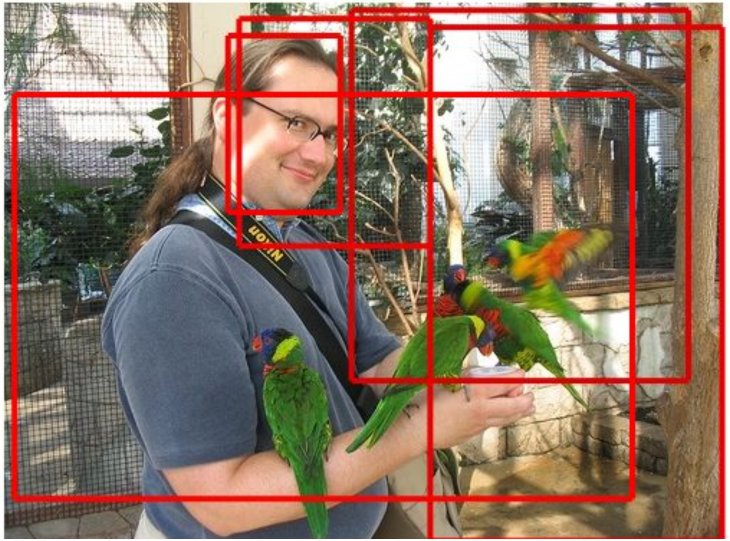
- Local, regional, and global features optimally combined through CRF $\vec{\lambda}^* = \arg\max_{\vec{\lambda}} \sum_n \log P(A^n | I^n; \vec{\lambda})$.
  - Elegant framework to combine multiple features
- Unliked MRF (GBVS), CRF's can use arbitrary features extracted from the **whole image**
- Exact computation of marginal distribution $p(a_x^n | I^n; \vec{\lambda})$ is intractable
- Jointly optimizing for 3 Salient Object Features $f_k(x, I) \in [0, 1]$:
  - Multi-scale contrast $f_c(x, I) = \sum_{l=1}^{L} \sum_{x' \in N(x)} \| I^l(x) - I^l(x') \|^2$
  - Center-surround histogram $f_h(x, I) \propto \sum_{\{x' | x \in R^*(x')\}} w_{xx'} \chi^2(R^*(x'), R_S^*(x'))$.
  - Color spatial-distribution $f_s(x, I) \propto \sum_c p(c | I_x) \cdot (1 - V(c)) \cdot (1 - D(c))$
- Bayesian Optimizable parameters:
  - **Learned Edge weights for Belief Propagation network (8 params)**
  - # clusters in color spatial distribution (1 param)
  - # levels of scale pyramid in Multi-scale contrast function (1 param)
  - Epsilon in weighted color spatial distribution (1 param)

Result: Objectness

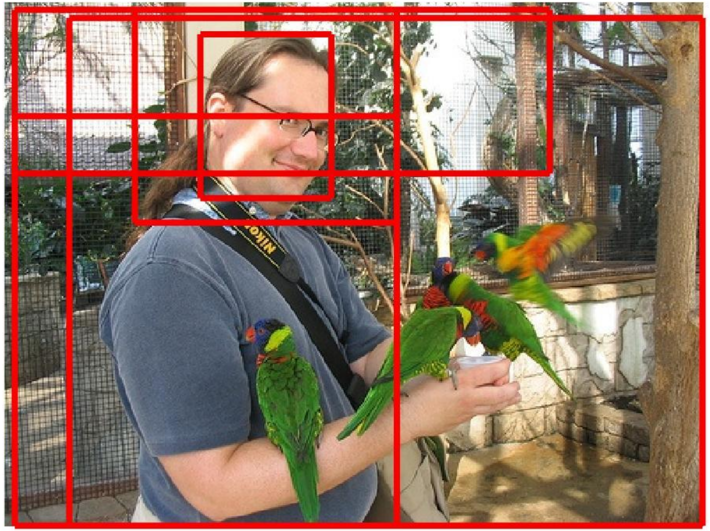
- 5 parameters that define multi-scale saliency score are optimized to maximize the total objectness score,with a perfect score being 5.
- 3 different algorithms, namely GP-UCB, GP-PI and GP-EI are compared in terms of their convergence rate and final score
- All 3 algorithms provided better scores compared to original parameters in the paper.
- GP-EI showed the best performance both in terms of convergence and the final score.



Comparison

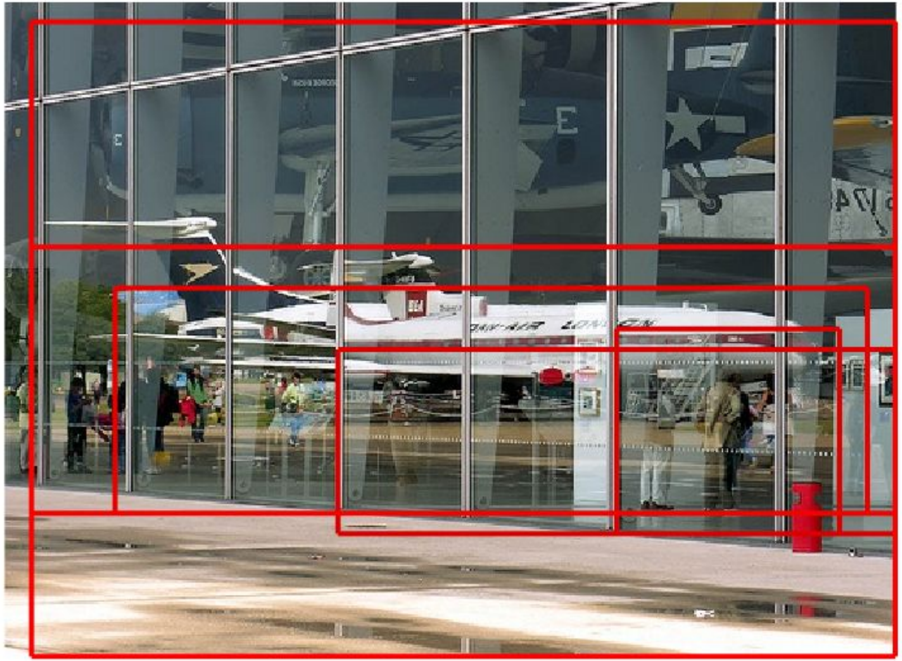


HAND-TUNED

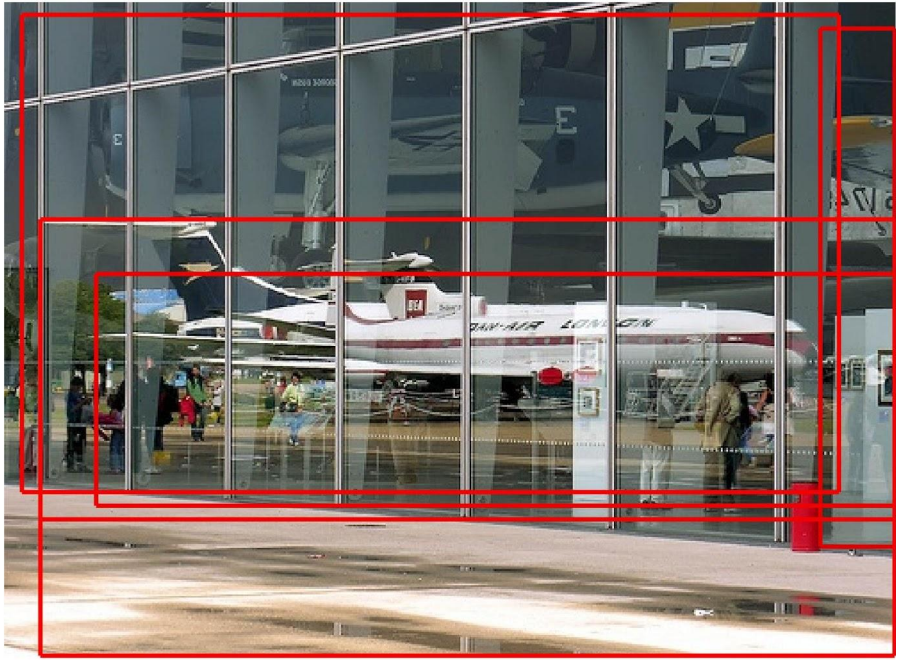


OPTIMIZED

Comparison



HAND-TUNED



OPTIMIZED

Comparison

HAND-TUNED

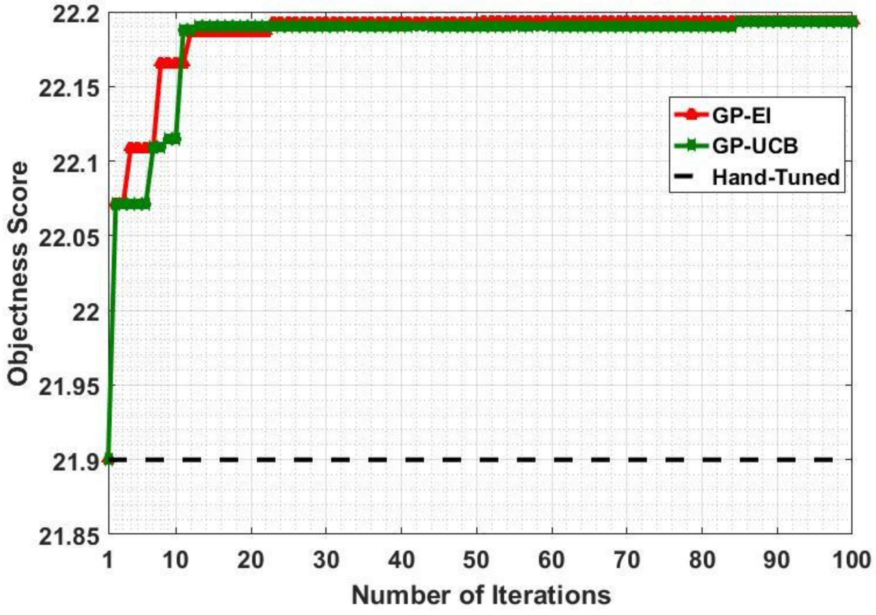
OPTIMIZED

Results: Saliency

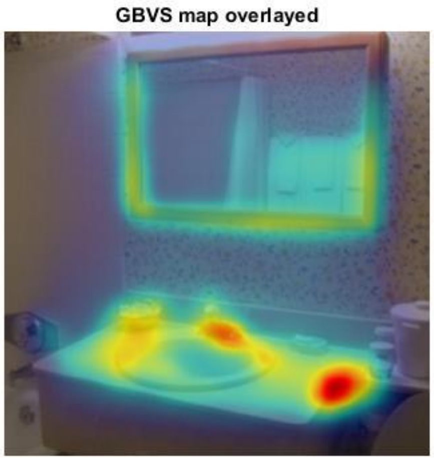
- ❏ 7 parameters that define weights of different channels in the image are optimized to maximize total saliency score. The total score is calculated for 30 different images, 30 being the perfect score.

- ❏ 2 different algorithms, namely GP-UCB and GP-EI are compared in terms of their convergence rate and final score.

- ❏ Both algorithms provided better scores compared to original parameters in the paper, which was equal weights from each channel.



Comparison: *"Graph-based Visual Saliency" vs. "Graph-based Visual Saliency" + BO*



HAND-TUNED

OPTIMIZED

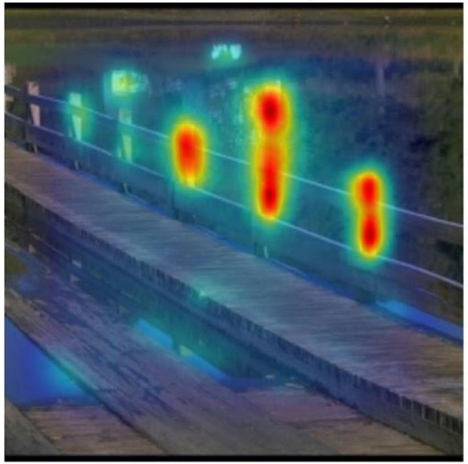
Comparison: *“Graph-based Visual Saliency” vs. “Graph-based Visual Saliency” + BO*

Original Image

GBVS map overlayed

HAND-TUNED

GBVS map overlayed



OPTIMIZED

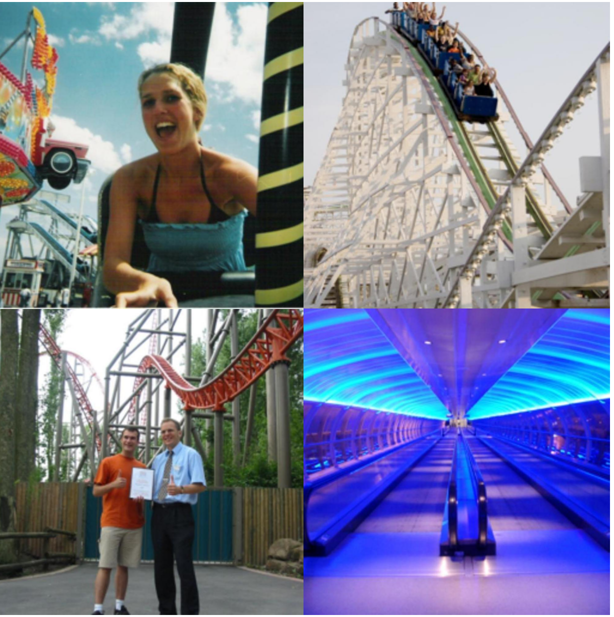
Comparison: *"Graph-based Visual Saliency" vs. "Graph-based Visual Saliency" + BO*

HAND-TUNED

OPTIMIZED

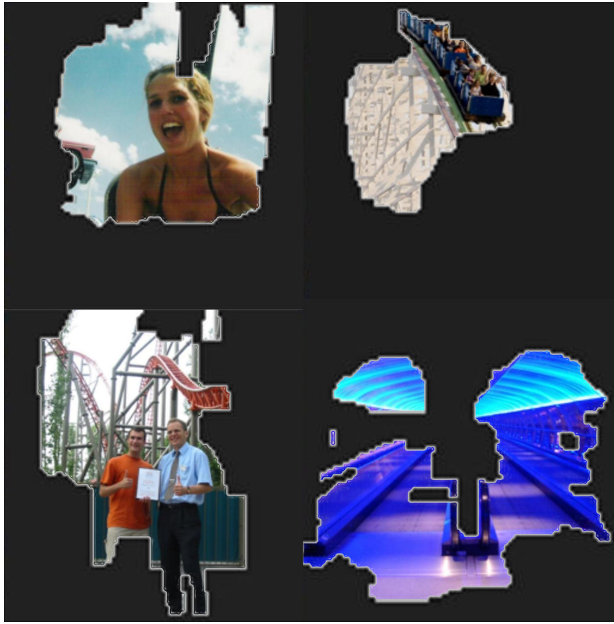
Comparison:

"Graph-based Visual Saliency" '07 + BayesOpt. vs. "Learning to Detect A Salient Object" '11



GBVS - optimized (ours)

MSRA - baseline



References

[1] B. Alexe, T. Deselaers, and V. Ferrari. What is an object? In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 73–80, June 2010.
[2] Gunhee Kim and Antonio Torralba. Unsupervised detection of regions of interest using iterative link analysis. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, Advances in Neural Information Processing Systems 22, pages 961–969. Curran Associates, Inc., 2009.
[3] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascalnetwork.org/challenges/VOC/voc2007/workshop /index.html.
[4] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In NIPS, 2006.
[5] T. Liu, J. Sun, N. N. Zheng, X. Tang and H. Y. Shum, "Learning to Detect A Salient Object," 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, 2007, pp. 1-8.